

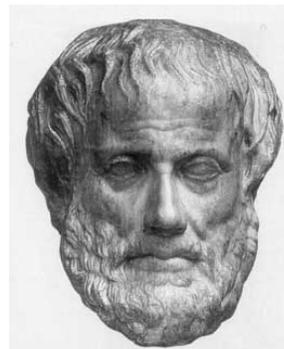
Ontology: A Vision for the Future and Its Realization

Mark A. Musen (Stanford University)
Barry Smith (University at Buffalo)



What Is An Ontology?

- The study of being
- A discipline co-opted by computer science to enable the explicit specification of
 - Entities
 - Properties and attributes of entities
 - Relations between entities
- A theory that provides a common vocabulary for an application domain



Porphyry's depiction of Aristotle's Categories

Supreme genus:

Differentiae:

Subordinate genera:

Differentiae:

Subordinate genera:

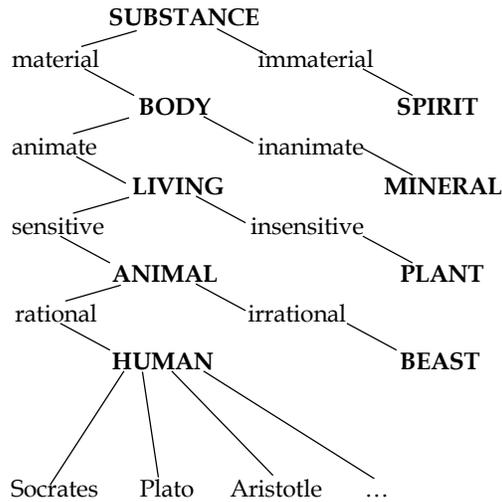
Differentiae:

Proximate genera:

Differentiae:

Species:

Individuals:



[Web](#)
[Images](#)
[Groups](#)
[Directory](#)
[News](#)

[Preferences](#)
[Directory Help](#)

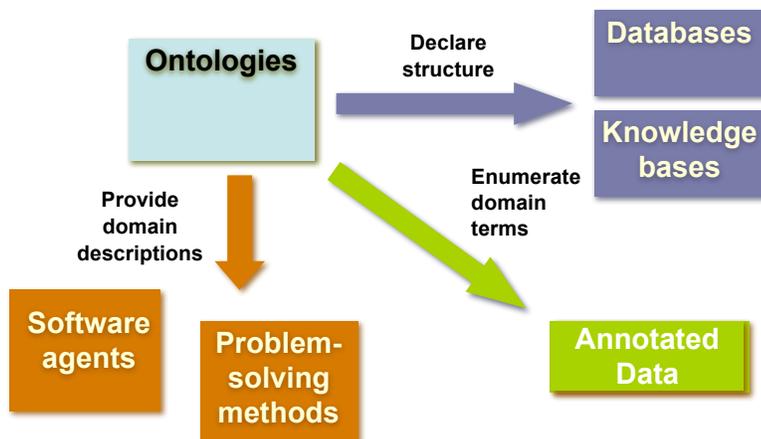
The web organized by topic into categories.

- | | | |
|---|--|---|
| <p>Arts
 Movies, Music, Television,...</p> | <p>Home
 Consumers, Homeowners, Family,...</p> | <p>Regional
 Asia, Europe, North America,...</p> |
| <p>Business
 Industries, Finance, Jobs,...</p> | <p>Kids and Teens
 Computers, Entertainment, School,...</p> | <p>Science
 Biology, Psychology, Physics,...</p> |
| <p>Computers
 Hardware, Internet, Software,...</p> | <p>News
 Media, Newspapers, Current Events,...</p> | <p>Shopping
 Autos, Clothing, Gifts,...</p> |
| <p>Games
 Board, Roleplaying, Video,...</p> | <p>Recreation
 Food, Outdoors, Travel,...</p> | <p>Society
 Issues, People, Religion,...</p> |
| <p>Health
 Alternative, Fitness, Medicine,...</p> | <p>Reference
 Education, Libraries, Maps,...</p> | <p>Sports
 Basketball, Football, Soccer,...</p> |
| <p>World
 Deutsch, Español, Français, Italiano, Japanese, Korean, Nederlands, Polska, Svenska, ...</p> | | |

Why develop an ontology?

- To share a **common understanding** of the entities in a given domain
 - among people
 - among software agents
 - between people and software
- To enable **reuse** of data and information
 - to avoid re-inventing the wheel
 - to introduce standards to allow interoperability and automatic reasoning
- To create **communities of researchers**

Ontologies are just the beginning



Ontologies are cropping up everywhere!

- Indexing of online information for access by humans or search engines
- Reference terminologies for machine translation and data interchange
- Standard terms for describing experimental data
- Frameworks for structuring knowledge for decision support



Srinija Srinivasan

ONTOLOGIST

She may be the best-kept secret at Yahoo!, the company that produces the wildly popular Web search engine. Trained in library science, Srinivasan is the one who decides how the thousands of Web pages submitted to Yahoo! should be categorized and classified, making it as intuitive, expandable and maintainable as possible.

42 NEWSWEEK DEC. 25, 1995/JAN. 1, 1996

Foundational Model of Anatomy

- Long-term project at University of Washington to create a comprehensive ontology of human anatomy
- 72K concepts, 1.9M relationships
- One of the largest and best developed ontologies in biomedicine



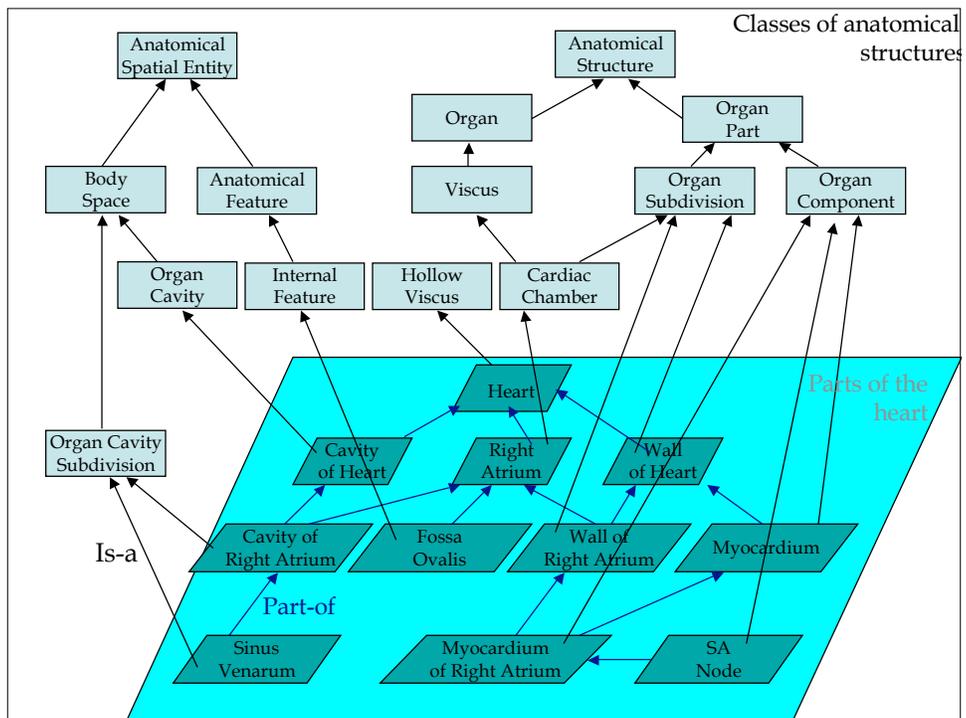
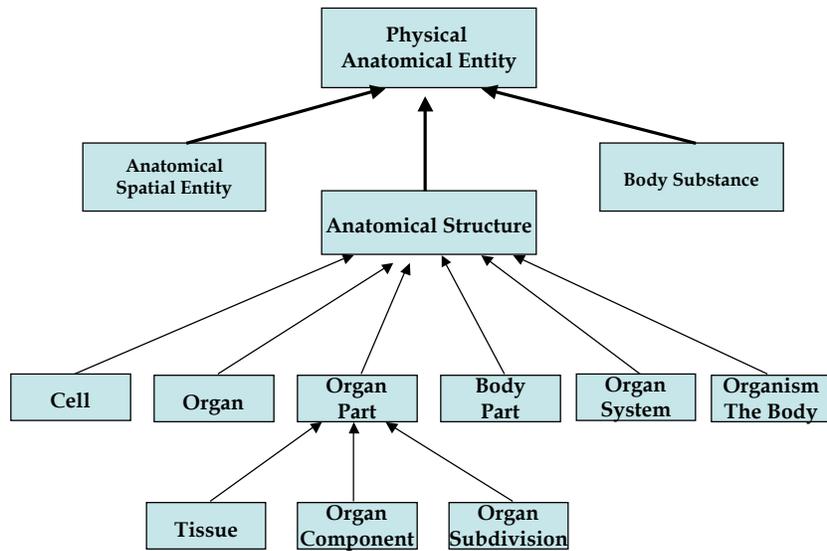
Structural Informatics Group
Structural Informatics Group
Structural Informatics Group

SIG

Overview Projects Personnel Publications Docs Local Info Products

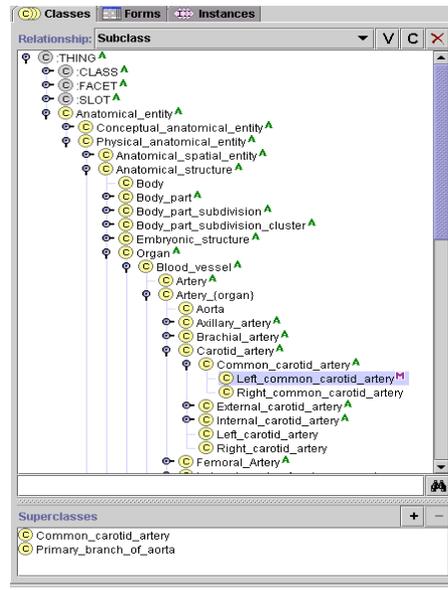
Biological Structure Biomedical Informatics Computer Science and Engineering University of Washington

Top level of the Foundational Model of Anatomy



We really want ontologies in electronic form

- Ontology contents can be processed and interpreted by computers
- Interactive tools can assist developers in ontology authoring



Web Images Groups **Directory** News

Google Search [Preferences](#) [Directory Help](#)

The web organized by topic into categories.

Arts Movies , Music , Television ,...	Home Consumers , Homeowners , Family ,...	Regional Asia , Europe , North America ,...
Business Industries , Finance , Jobs ,...	Kids and Teens Computers , Entertainment , School ,...	Science Biology , Psychology , Physics ,...
Computers Hardware , Internet , Software ,...	News Media , Newspapers , Current Events ,...	Shopping Autos , Clothing , Gifts ,...
Games Board , Roleplaying , Video ,...	Recreation Food , Outdoors , Travel ,...	Society Issues , People , Religion ,...
Health Alternative , Fitness , Medicine ,...	Reference Education , Libraries , Maps ,...	Sports Basketball , Football , Soccer ,...
World Deutsch , Español , Français , Italiano , Japanese , Korean , Nederlands , Polska , Svenska , ...		

Open Directory Project

- Started in 1998 as a volunteer effort to develop an open-content directory of Web pages
- In its first year, 4500 editors had indexed 100K Web sites
- By July 2005, 69K editors had indexed 4.6M sites using 580K categories
- On average, between 9K and 10K volunteer editors are working on ODP at any given time

 open directory project

The New Philosophers

- Categorizing what exists in machine-understandable form
- Providing a structure that enables
 - developers to locate and update relevant descriptions
 - computers to infer relationships and properties
 - quantitative data to be annotated in such a way as to become available for semantically meaningful search

Lots of ontology builders are not very good philosophers

- Nearly always, ontologies are created to address pressing practical needs
- The people who have the most insight into professional knowledge of a given biomedical domain may have little appreciation for metaphysics, principles of knowledge representation, or computational logic
- There simply aren't enough good philosophers to go around

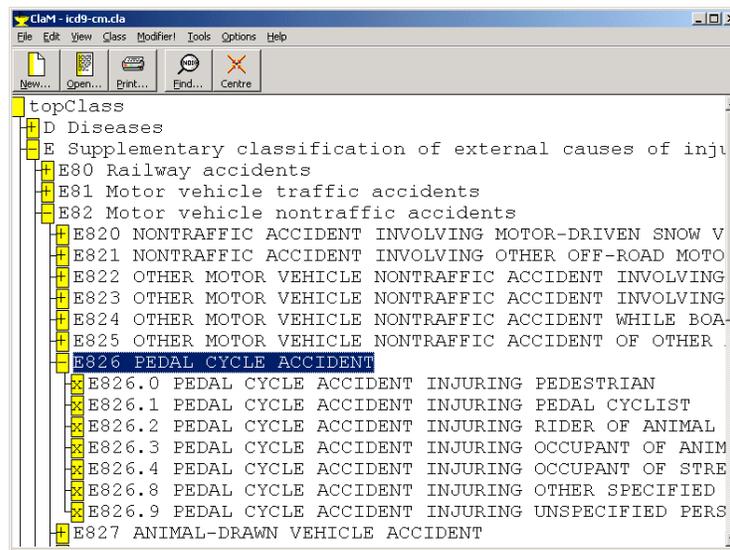
A case in point: The International Classification of Diseases

- An enumeration of diseases that forms the basis for all medical claims and reimbursements in the United States
- A “legacy” terminology that has its roots in 19th century epidemiology
- Created initially by biostatisticians with a pressing need to compare death statistics in different European countries
- A system that won't go away—and yet we would never create anything like it again

The International Classification of Diseases

724 Unspecified disorders of the back
724.0 Spinal stenosis, other than cervical
724.00 Spinal stenosis, unspecified region
724.01 Spinal stenosis, thoracic region
724.02 Spinal stenosis, lumbar region
724.09 Spinal stenosis, other
724.1 Pain in thoracic spine
724.2 Lumbago
724.3 Sciatica
724.4 Thoracic or lumbosacral neuritis
724.5 Backache, unspecified
724.6 Disorders of sacrum
724.7 Disorders of coccyx
724.70 Unspecified disorder of coccyx
724.71 Hypermobility of coccyx
724.71 Coccygodynia
724.8 Other symptoms referable to back
724.9 Other unspecified back disorders

ICD9 (1977): A Handful of Codes for Traffic Accidents



ICD10 (1999): 587 codes for such accidents

<p>V31.22 Occupant of three-wheeled motor vehicle injured in collision with pedal cycle, person on outside of vehicle, nontraffic accident, while working for income</p>	<p>W65.40 Drowning and submersion while in bath-tub, street and highway, while engaged in sports activity</p>	<p>X35.44 Victim of volcanic eruption, street and highway, while resting, sleeping, eating or engaging in other vital activities</p>
--	---	--

ICD is used for lots of (too many?) things!

- ICD is used to code all patient encounters with the health-care system for purposes of
 - Billing and reimbursement
 - Institutional planning
 - Disease surveillance and public health
 - Quality assurance
 - Economic modeling by third-party payors
- ICD was never intended to make the distinctions relevant to all these tasks!

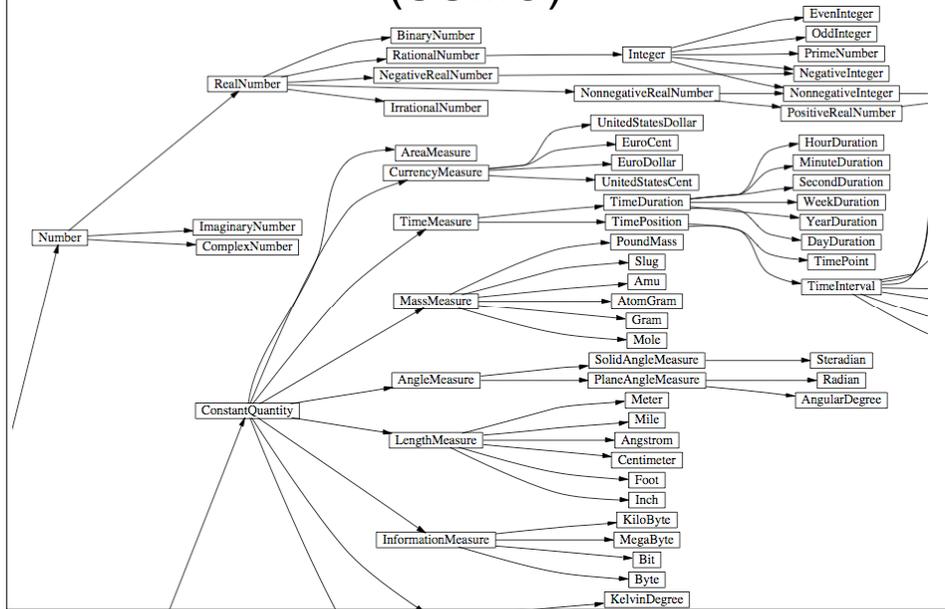
If real ontologists could build the ICD from scratch ...

- Diseases would be organized with well-defined relationships
- Diseases would be associated with computer-understandable definitions
- There would be well-defined rules for ensuring that descriptions are sensible
- There would be well-defined mechanisms for creating use-specific views of the ICD
- There would be a well-defined path to integration with bioinformatics resources that describe the molecular underpinnings of disease

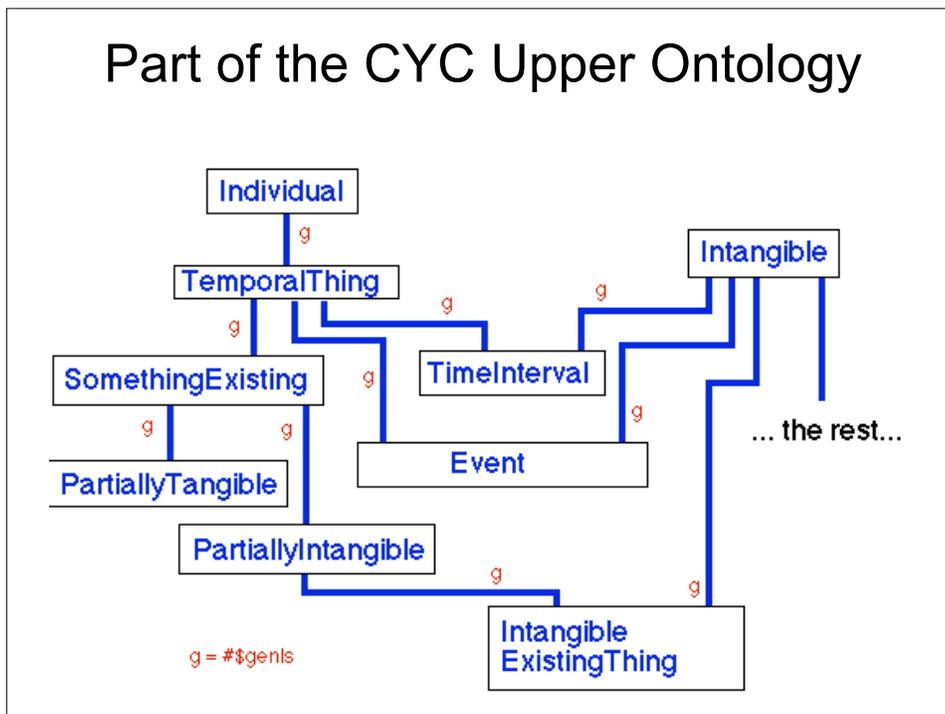
Distinctions among ontologies

- **“Upper-level” versus “domain-oriented”**: Does the ontology try to describe general, abstract entities or entities tied to a particular application area?
- **“Light” versus “heavy”**: Is the ontology a simple taxonomy or does the ontology provide additional detail regarding the nature of entities?

Suggested Upper Merged Ontology (SUMO)



Part of the CYC Upper Ontology



Some CYC Definitions

#\$Cancer

isa: #PhysiologicalConditionType
genls: #AilmentCondition
#TerminalPhysiological Condition

#\$Tumor

isa: #ExistingObjectType
genls: #BiologicalLivingObject

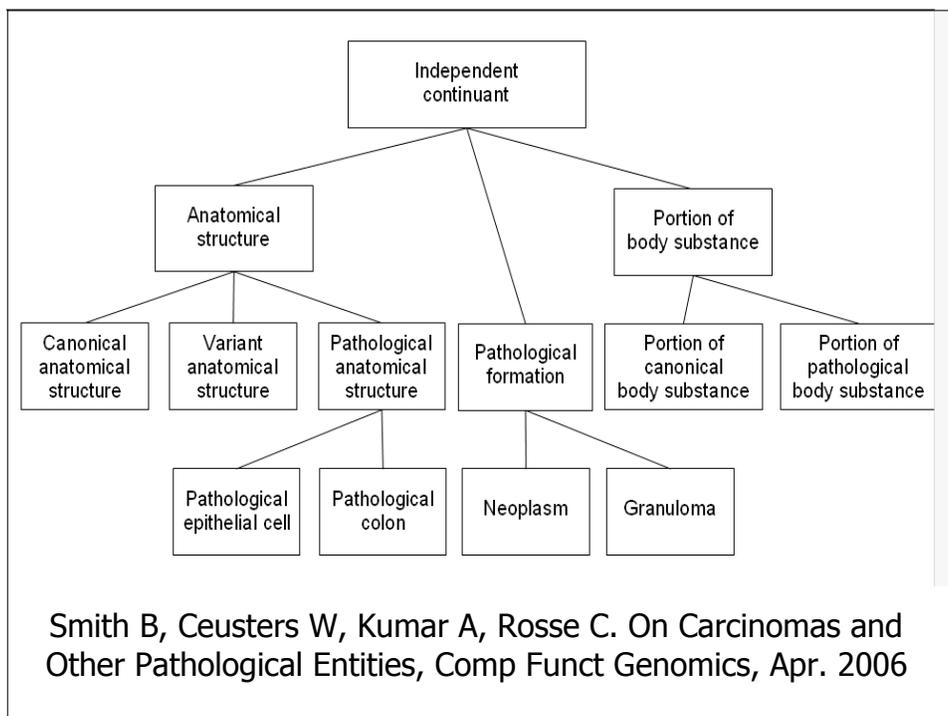
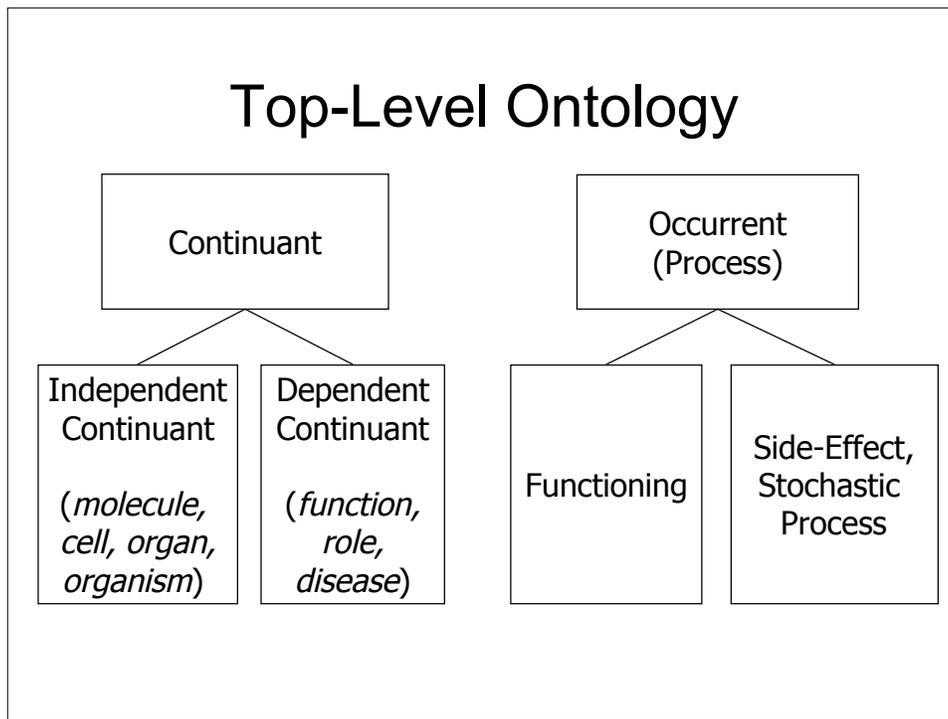
#\$Infection

isa: #PhysiologicalConditionType
genls: #AilmentCondition

Basic Formal Ontology

An example of a top-level ontology being used in biomedicine to structure ontologies for specific domains that are marked by internal coherence and external interoperability

Top-Level Ontology



Distinctions among ontologies

- **“Upper-level” versus “domain-oriented”**: Does the ontology try to describe general, abstract entities or entities tied to a particular application area?
- **“Light” versus “heavy”**: Is the ontology a simple taxonomy or does the ontology provide additional detail regarding the nature of entities?

Taxonomies are “Light-Weight” Ontologies

724 Unspecified disorders of the back
724.0 Spinal stenosis, other than cervical
724.00 Spinal stenosis, unspecified region
724.01 Spinal stenosis, thoracic region
724.02 Spinal stenosis, lumbar region
724.09 Spinal stenosis, other
724.1 Pain in thoracic spine
724.2 Lumbago
724.3 Sciatica
724.4 Thoracic or lumbosacral neuritis
724.5 Backache, unspecified
724.6 Disorders of sacrum
724.7 Disorders of coccyx
724.70 Unspecified disorder of coccyx
724.71 Hypermobility of coccyx
724.71 Coccygodynia
724.8 Other symptoms referable to back
724.9 Other unspecified back disorders

Properties of a class (e.g., “Esophagus”)

CLASS EDITOR

For Class: Esophagus (instance of Organ with organ cavity)

Preferred Name

Esophagus

Has Boundary Has Dimension Has Mass Has Inherent 3-D Sh...

Dimension: 3-dimension

Inherent 3-D Shape:

Hollow cylinder

Has Physical State

Solid

Physical State

Regional Part

- Cervical part of esophagus
- Thoracic part of esophagus
- Abdominal part of esophagus

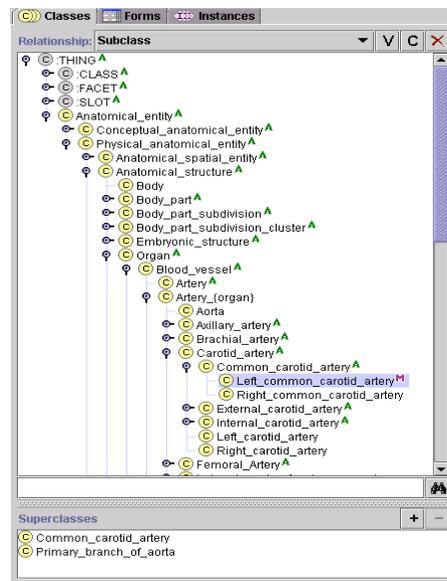
Regional Part Of

- Foregut
- Upper gastrointestinal tract
- Gut
- Gastrointestinal system

is-a is a special relation

If a sub-class *is-a* member of a super-class, then

- every instance of the sub-class is also an instance of the super-class (e.g., every member of the set *aorta* is necessarily a member of the set *artery*)
- values of attributes of the super-class are *inherited* by every instance of the sub-class (e.g., if *arteries* have cylindrical shape, then *aorta* has cylindrical shape)



Modeling *part-of* relationships is tricky

- Inheritance is not necessarily transitive
 - In an *is-a* relation, if a stomach is an organ and an organ has a volume, then a stomach has volume
 - In a *part-of* relation, if an eyebrow is part of the head and the head has a volume, then does an eyebrow have a volume?
- There are many kinds of *part-of* relationships, each with slightly different semantics

Kinds of *part-of* relationships (after Winston and Odell)

- Component (e.g., handle of a car door)
- Stuff (e.g., flour in bread)
- Portion (e.g., a slice from a loaf of bread)
- Area (e.g., city in a country)
- Member (e.g., ship in a fleet of ships)
- Partner (e.g., Laurel in Laurel & Hardy)
- Piece (e.g., handle when removed from the door)

“Frame-based” knowledge-representation systems

- Allow developers to encode
 - Taxonomic hierarchies of classes
 - Other relations among classes (e.g., *part-of*) in addition to the *is-a* hierarchy
 - Attributes of classes that take on particular values to define *instances* of the classes
- Support inheritance of attributes and values along taxonomic relations

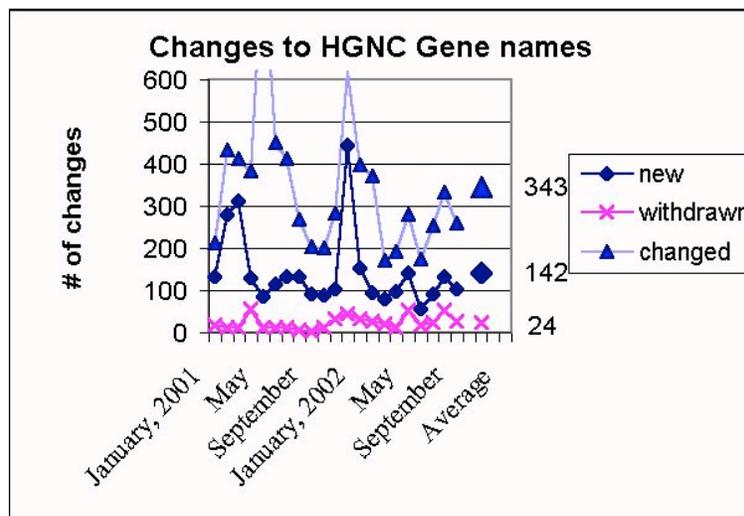
The story so far ...

- Ontologies define the entities—and relationships among entities—in some application area
- The authors’ point of view determines which distinctions are appropriate in a particular ontology
- Ontologies often use frame-based representations (including classes, attributes, relationships, and axioms) to encode knowledge
- People are building ontologies for nearly every niche of biomedicine

A Parable: The Gene Ontology



The pressing need to standardize the names of human genes



But the human genome is only part of the problem ...

- Biologists maintain huge databases of gene sequences and gene expression for a wide range of “model organisms” (e.g., mouse, rat, yeast, fruit fly, round worm, slime mold)
- Database entries are annotated with entries such as the name of a gene, the function of the gene, and so on
- How do you ensure uniformity of these annotations?

Gene Ontology Consortium

- Founded in 1998 as a collaboration among scientists responsible for developing different databases of genomic data for model organisms (fruit fly, yeast, mouse)
- Now, essentially all developers of all model-organism databases participate
- Goal: To produce a dynamic, controlled vocabulary that can be applied to all organism databases even as knowledge of gene and protein roles in cells is accumulating and changing



Gene Ontology (GO)

- Comprises three independent “ontologies”
 - **molecular function** of gene products
 - **cellular component** of gene products
 - **biological process** representing the gene product’s higher order role.
- Using these terms to express attributes of gene products in the collaborating databases allows queries across databases, providing linkage of biological information across species



AmiGO Search GO:

Terms Gene Products

[Top Docs](#) [Gene Ontology](#) [GO Links](#) [GO Summary](#)

☐ **GO:0003673 : Gene Ontology (92932)**

- ☐ **GO:0008150 : biological process (56952)**
 - ☐ GO:0007610 : behavior (566)
 - GO:0000004 : biological process unknown (6152)
 - ☐ **GO:0007154 : cell communication (11916)**
 - ☐ GO:0007155 : cell adhesion (830)
 - GO:0030260 : cell invasion (0)
 - ☐ GO:0008037 : cell recognition (210)
 - ☐ **GO:0007267 : cell-cell signaling (1318)**
 - ☐ GO:0045168 : cell-cell signaling involved in cell fate commitment (0)
 - ☐ **GO:0030072 : peptide hormone secretion (6)**
 - GO:0030252 : growth hormone secretion (2)
 - GO:0030073 : insulin secretion (4)
 - GO:0030103 : vasopressin secretion (2)
 - ☐ GO:0019226 : transmission of nerve impulse (688)
 - ☐ GO:0030383 : host-pathogen interaction (12)

GO is wildly successful

- Biologists around the world contribute to GO on a regular basis
- The ontology is updated every 30 minutes!
- It's now impossible to work in most areas of computational biology without making use of GO terms

GO has faced real problems ...

- Ontologies use an idiosyncratic format not compatible with standard knowledge-representation systems
- Because of the informal knowledge-representation system, errors crept into GO
 - terms duplicated in different places
 - terms with no superclasses
 - uncertain relationships between terms
- The GO Consortium has embraced an ambitious strategy to fix these problems
 - new representation system (OBO-Edit)
 - standardized relationships among entities
 - enhanced quality control
 - more sophisticated Web-based tools: NCBO's BioPortal

Creating ontologies has become a widespread cottage industry

- Professional Societies
 - MGED: Microarray Gene Expression Data Society Ontology
 - HUPO: Human Protein Organization Ontology
- Government
 - NCI Thesaurus
 - NIST: Process Specification Language
- Open Biomedical Ontologies
 - GO
 - Three dozen (and growing) other biomedical ontologies



A Portion of the OBO Library

Domain	Prefix	Ontology	Defs file
Arabidopsis gross anatomy	TAIR	arabidopsis anatomy.ontology	arabidopsis anatomy.definitions
Arabidopsis development	TAIR	arabidopsis development.ontology	arabidopsis development.definitions
Cell type	CL	cell.obo	Included in cell.obo
Cereal plant gross anatomy	GRO	anatomy gr ont	anatomy gr def
Cereal plant development	GRO	temporal gr ont	temporal gr def
Cereal plant trait ontology	TO	trait ontology	trait definitions
Chemical entities of biological interest	CHEBI	ontology.obo	Included in ontology.obo
Protein covalent bond	CV	[none]	[none]
Protein-protein interaction	MI	psi-mi.dag	psi-mi.def
Maize gross anatomy	ZEA	Zea mays anatomy ontology.txt	Zea mays anatomy ontology definitions.txt
Dictyostelium anatomy	DDANAT	anatomy.ontology	anatomy.definitions
Drosophila gross anatomy	FBbt	fly anatomy.ontology	fly anatomy.definitions
Habronattus courtship		protege source	Included in protege source
Loggerhead nesting		protege source	Included in protege source
Human anatomy and development	EV	ontologies	[none]
Microarray experimental conditions		MGEDOntology.daml	Included in MGEDOntology.daml
Physical-chemical methods and properties	FIX	fix.ontology	[none]
Fungal gross anatomy	FAO	fungal anatomy.ontology	fungal anatomy.definitions
Molecular function	GO	gene_ontology.obo	Included in gene_ontology.obo
Biological process	GO	gene_ontology.obo	Included in gene_ontology.obo
Cellular component	GO	gene_ontology.obo	Included in gene_ontology.obo

Ontologies are meeting an urgent need

- Ontologies are being developed by interested groups from every sector of academia, industry, and government
- Many of these ontologies have been proven to be extraordinarily useful to wide communities
- We finally have tools and representation languages that can enable us to create durable and maintainable ontologies with rich semantic content

The National Center for Biomedical Ontology

- One of three National Centers for Biomedical Computing launched by NIH in 2005
- Collaboration of Stanford, Berkeley, Mayo, Buffalo, Victoria, UCSF, Oregon, and Cambridge
- Primary goal is to make ontologies accessible and usable
- Research will develop technologies for ontology indexing, alignment, and peer review



NATIONAL CENTER FOR BIOMEDICAL ONTOLOGY *BioPortal*

Browse Search Align

Ontologies

View By

Select item and ...

[Expand All](#) | [Collapse All](#)

Ontologies

Select	Focus	Ontology	Knowledge Representation Language	Foundry	Current Version	Release Date	Version Status
<input type="radio"/>		Ontologies					
<input type="radio"/>	<input checked="" type="checkbox"/>	Anatomy					
<input type="radio"/>		BRENDA tissue / enzyme source	OBO Text	No	1.1	5/3/2006	Production
<input type="radio"/>		Cell type	OBO Text	Yes	1.1	5/3/2006	Pre-Production
<input type="radio"/>		Drosophila gross anatomy	OBO Text	Yes	1.1	5/3/2006	Production
<input type="radio"/>		Mesquite gross anatomy	OBO Text	No	1.1	5/3/2006	Production
<input type="radio"/>	<input checked="" type="checkbox"/>	Gross Anatomy					
<input type="radio"/>	<input checked="" type="checkbox"/>	Animal Gross Anatomy					
<input type="radio"/>	<input checked="" type="checkbox"/>	Fish Anatomy					
<input type="radio"/>		Medaka fish anatomy and development	OBO Text	Yes	1.1	5/3/2006	Pre-Production
<input type="radio"/>		Zebrafish anatomy and development	OBO Text	Yes	1.1	5/3/2006	Production
<input type="radio"/>	<input checked="" type="checkbox"/>	Human Developmental Anatomy					
<input type="radio"/>		View details of the current ontology.					
<input type="radio"/>		Human developmental anatomy, abstract version	OBO Text	Yes	1.1	5/3/2006	Production
<input type="radio"/>		Human developmental anatomy, timed version	OBO Text	Yes	1.1	5/3/2006	Production
<input type="radio"/>	<input checked="" type="checkbox"/>	Mouse Anatomy					
<input type="radio"/>		Mouse gross anatomy and development	OBO Text	Yes	1.1	5/3/2006	Production
<input type="radio"/>		Mouse adult gross anatomy	OBO Text	Yes	1.1	5/3/2006	Production
<input type="radio"/>	<input checked="" type="checkbox"/>	Microbial Anatomy					
<input type="radio"/>		Dictyostelium discoideum anatomy	OBO Text	Yes	1.1	5/3/2006	Pre-Production
<input type="radio"/>		Fungal gross anatomy	OBO Text	Yes	1.1	5/3/2006	Production
<input type="radio"/>	<input checked="" type="checkbox"/>	Plant Anatomy					
<input type="radio"/>		Maize gross anatomy	OBO Text	Yes	1.1	5/3/2006	Pre-Production
<input type="radio"/>		Cereal plant gross anatomy	OBO Text	Yes	1.1	5/3/2006	Retired
<input type="radio"/>	<input checked="" type="checkbox"/>	Chemical					

NATIONAL CENTER FOR BIOMEDICAL ONTOLOGY *BioPortal*

Browse Search Align

Logged in as test | [My Account](#) [Sign Out](#)

Project Information File and Version Foundry Review

Submit New Ontology - Project Information

Step 1 of 4

• Indicates required field

[Hide Optional Fields](#)

• Display Name

• Knowledge Representation Language

• Categories

• Supported Language

Description

Contact Name

Contact Email

Home Page

Documentation Page

Publications Page

Step 1 of 4

Goals for BioPortal

- Web accessible repository of ontologies for the biomedical community
- Support for ontology
 - Peer review
 - Annotation (marginalia)
 - Versioning
 - Alignment
 - Search



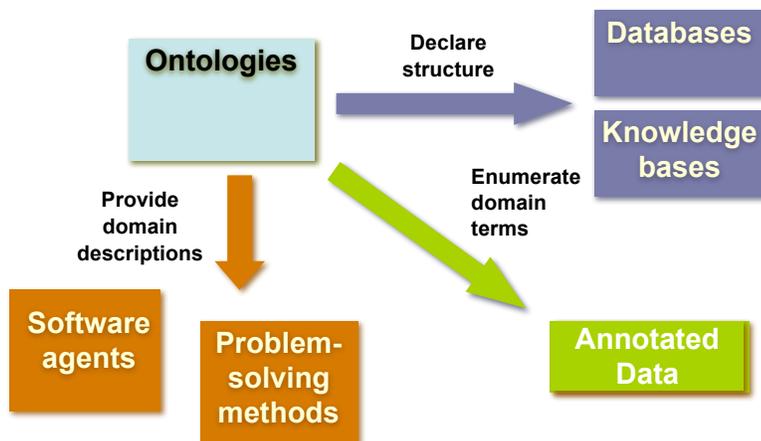
Other Center Activities

- Biological Driving Projects that will use BioPortal ontologies to annotate biomedical data
- Collaborating projects that will use BioPortal ontologies for
 - natural-language processing
 - information integration
 - data and knowledge visualization
- Outreach activities to help different communities to build better ontologies and to utilize the Center's technology

A thousand flowers are blooming!

- Ontologies are being developed by interested groups from every sector of academia, industry, and government
- Many of these ontologies have been proven to be extraordinarily useful to wide communities
- We finally have tools and representation languages that can enable us to create durable and maintainable ontologies with rich semantic content

Ontologies are just the beginning



Our Center will offer

- Technology for uploading, browsing, and using biomedical ontologies
- Methods to make the online “publication” of ontologies more like that of journal articles
- Tools to enable the biomedical community to put ontologies to work on a daily basis

